IJCSE International Journal of Computer Sciences & Engineering Open Access
Review Paper Volume-4, Special Issue-3, May 2016 E-ISSN: 2347-2693

Architecture for remote intelligent data processing

(secure, scalable and versatile multi-layer client–server)

Manjunath R[1] and Shivarayappa Maranur[2]

[1,2] Dept. of Computer Science & Engineering City Engineering College Bangalore, India

Available online at: www.ijcseonline.org

***Abstract*—** In recent years, the need for data collection and Analysis is growing in many scientific disciplines. This is Consequently causing an increase of research in automated data management and data mining to create reliable methods for data analysis. To deal with the need for smart environments and big computational resources, some previous works proposed to address the problem by moving on remote processing, with the aim of sharing supercomputer resources, algorithms and costs. Following this trend, in this work we propose an architecture for advanced remote data processing in a secure, smart and versatile client–server environment that is capable of integrating pre-existing local software. In order to assess the feasibility of our proposal, we developed a case study in the context of an image-based medical diagnostic environment. Our tests demonstrated that the proposed architecture has several benefits: increase of the system throughput, easy upgradability, maintainability and scalability. Moreover, for the scenario we have considered, the system showed a very low transmission overhead which settles on about 2.5%for the widespread 10/100 mbps.

Keywords *– JAAS,DCE-MRI ,OsiriX, secure ,NIST , biomedical ,Image processing ,TLS/SSL*

I.INTRODUCTION

As technology advances at a rapid pace, data collection and analysis is becoming more important. Research on automated data management (storage and acquisition) and data mining has increased with the aim to create more reliable methods [1]. For example, in retail business, Wal-Mart Stores, Inc. handles more than 1 million customer transactions every hour. A data warehouse estimated to contain more than 3 PB of data shows record of every single purchase by their point of-sale terminals in each of their 6000 stores worldwide. By applying machine learning on this data, they can extract patterns indicating the effectiveness of their pricing and advertising strategies ensuring a wide knowledge for better inventory management, supply, and promotion and advertising campaign [2]. Medical diagnostic imaging usually requires massive processing of huge data, often with strong temporal deadline constrains. For example, a small clinical center equipped for Magnetic Resonance Imaging (MRI) can produce up to 20–25 MRI scans per day, resulting in about 4 GB of raw data per day, which can easily tenfold after processing. Such amount of data is likely to quickly grow because of diagnostic imaging technologies improvements. Local workstation cannot handle resource-consuming processing on big dataset without involving a large amount of time and vice versa due to the limited hardware capabilities. The constraints for medical image processing in small and medium companies.

II. RELATED WORKS

Some studies proposed very general purpose architecture for remote processing in different fields (domain-independent Frameworks). Triana [3] is a visual workflow-based problem solving software, developed at CardiffUniversity.Originally, it was developed for a gravitational wave detection project.

Subsequently, it has been extended to incorporate a range of modules, such as peer-to-peer communication, grid services and web services integration with the purpose of providing a wider integration with existing grid technologies [4]. The Kepler Project [5] is a scientific project that offers construction, composition, and orchestration engine. The focus is on data analysis and modelling, which influenced the design in that it is suitable for modelling processes in a wide variety of scientific domains from physics via ecosystems, to bioinformatics web services. Instead of trying to provide a generic semantic for all possible types of processes encountered in these domains, Kepler separates the execution engine from the workflow model and assigns one model of computation to each workflow [4]. The LONI Pipeline Processing Environment [6] is a java-based modules oriented visual programming interface for the simplified design and execution of remote pipelines. It was initially proposed to investigate for image processing in brain mapping. The Tavern Suite [7] is a powerful scientific workflow management system born to satisfy the needs of bioinformaticians who need to build scientific workflows from numerous remote web services.

It offers a graphical designing tool, an authoring client,

a workflow representation language and different additional components such as a service directory, data and meta-data repositories and others

III. THE PROPOSED ARCHITECTURE

In this paper, we propose a generic architecture for remote processing of data, with a special attention to security and scalability issues. Moreover, a preliminary implementation of our proposal has been realized using low-impact technologies and up-to-date standards.

We designed the system with the following three constraints in mind:

– Easy integration into third-party client-side software;

– TCP/IP network infrastructure;

– Up-to-date open standards.

A. Architecture layers

The system is organized as a multilayer client–server architecture implementing a multi-client/single-server model. Within the same node, different layers communicate using the file system (see Fig. 1).

Each layer has a well-defined role:

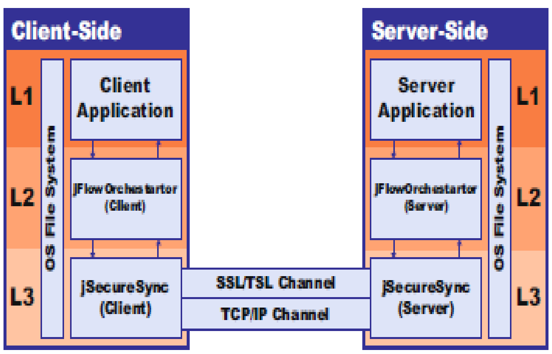

Fig. 1 Layers overview of the proposed architecture that provides: communication (L1), orchestration (L2) and application (L3)

B. Communication layer

The component which implements connection establishment and management, command transmission and file-state synchronization is called jSecureSync. It has been coded in Java in order to increase portability.

Running both on client-side and on server-side, it implements the multi-client/single-server model: the client starts communication, then the server creates a new thread for each accepted incoming connection. Once communication is established, client and server-side operate symmetrically: on both sides jSecureSync runs a Connection Manager module that manages every aspect of file synchronization over the opened channel. After the Connection Manager starts, client and server act like peers.

Client and server modules establish two different channels (Fig. 1) exploiting Java TCP/IP sockets:

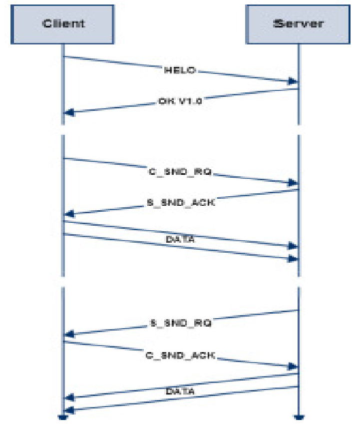

Fig. 2 Client–server communication protocol timing

Clients and server communicate through a protocol with low overhead (coded in hexadecimal format). Figure 2 presents the protocol commands:

C. Orchestrator layer

Object (jFlowOrchestrator) that provides all abstract methods to perform brokerage between L3 and L1, by orchestrating the steps of each request and the server-side computational flow progress.

It also provides adaptation for the application layer (L3) data inputs and outputs: different software can run on the application level both on server and client-side requiring different file format or data arrangement.

For example, as in the case study we will present in another section, medical image software could require a remote processing of a medical study. However, it is not required to send all the data to the CAD server (e.g. private information concerning patient name, age, etc. must be dealt with separately).

Exploiting the underlying layer (L1) services, the Orchestrator layer (L2) takes charge over the adaptation and coordination of the data flows. The jFlowOrchestrator provides three main abstract methods:

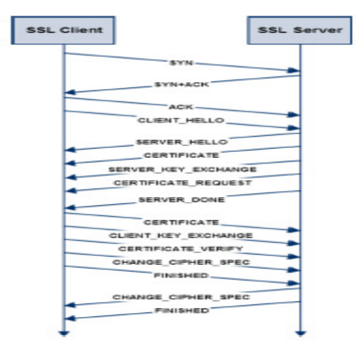

Fig. 3 SSL client-authenticated mode timing protocol

D. Application layer

The top layer (L3) has to be implemented according to client side software and server-side advanced processing software and tools. Client and server-side software must provide away to implement a communication with L2 layer (over file system) to:

– request a job (client);

– serve a job request (server);

– execute progress synchronization (client and server);

– provide results (server);

– retrieve results (client).

IV. A CASE STUDY

Dynamic Contrast Enhanced Magnetic Resonance Imaging (DCE-MRI) is a widespread methodology for cancer detection and evaluation. In particular, in this study, we focused on the breast DCE-MRI which has a large potential in early detection of cancer and therapy response assessment, especially for young women. Due to the large amount of data acquired within a DCEMRI session, the physician has to deal with the complex task of accurately detecting suspicious ROIs and evaluating tumor aggressiveness using 4D data (3D spatial data and 1D time data) [24,25]. In this context, many scenarios could be imagined in which the proposed architecture could be useful: let us consider the processing of multiple data corresponding to the same patient (patient follow-up) or the concurrent processing from multiple radiologists within a single medical center. We now describe the implementation of our proposal realized to support this application context.

V. CASE STUDY RESULTS

Benefits consist of both business and clinical areas. In today’s world of cost cutting, many facilities must show clinical benefit in order to justify expenditures, and the cloud technologies are potential tools to do just that.

1) Patient dataset

The single biggest clinical benefit that cloud technology can provide is access to applications that were previously unattainable. For example, the implementation of digital pathology, managed through cloud services, has a huge clinical impact on an organization. The organization can roll out a service that would have cost millions just for the storage alone, but now can pay for it as they use it. Access to pathologists who previously were reachable only near centre of excellence means that remote facilities can offer new services to the local patient population, relying on remote experts to render their diagnoses. Patient care can be improved by providing this service through the cloud faster and more efficiently. Since patients do not need to travel, waiting lists are more easily managed as more patients can have the same tests in more locations with a larger availability of experts. These same experts can access patient data remotely and on demand through the Internet via a 0ariety of connected devices. Physicians can review the latest diagnostic results from home and perhaps determine that the patient can be discharged immediately, rather than wait for their afternoon rounds.

2) Compression tests

Obviously there must be some business benefit for a new technology to be adopted, or it won't be considered. Cloud technologies provide tremendous benefits that can contribute to the welfare of a provider organization. Healthcare providers are in the business of treating and caring for patients. They are not IT focused; their purchasing patterns indicate that investment into IT falls far below other industry standards. In many cases providers IT staffs are stretched very thin, and other staff must assist. For example, in radiology it is often a medical technologist with a technical affinity but no formal technical background, who becomes the PACS administrator. The cloud offers providers the ability to access specific experts to manage and maintain their systems. A cloud provider will have a block storage expert, a network security expert, and an archiving and backup expert who will manage the different components. Providers need not build up these skill sets, but instead can, for example, focus on a clinical applications specialist for PACS who helps clinical users maximize the application. These experts can spend the time and effort to implement the best practices for each component, which ultimately delivers added benefit to the clinical users and their patients. Obviously there must be some business benefit for a new technology to be adopted, or it won't be considered. Cloud technologies provide tremendous benefits that can contribute to the welfare of a provider organization.

VI. DISCUSSION AND CONCLUSIONS

Healthcare providers face many challenges as they investigate moving to a cloud model.

A. Privacy challenges

Privacy and security rank at the top of the list of reasons for slow adoption rates. Putting personal health information into a 3rd-party, remote data center raises red flags where patient privacy laws are concerned. The possibility that patient data could be lost, misused or fall into the wrong hands affects adoption. What recourse does an organization have should a cloud provider lose data? It has happened, and it has the potential to be a very expensive problem to resolve. Violation of patient confidentiality carries heavy fines, including significant costs of recovery and patient notification. A potential solution is a private cloud model. In this case the data still resides at the customer data center and a certain degree of control still exists for organizations to manage patient privacy. This model may be more expensive, but security and privacy are more visible.

B. *Security challenges*

This may be a most important point where healthcare providers are concerned. One of the benefits of cloud technology is the ability to access resources that would otherwise be unattainable. A cloud provider will have security experts deploying the latest patches and software to its data center. Secure access to the physical property will be well guarded, and many policies, processes and mechanisms will be in place to ensure data security. Add to that the fact that any applications operating through the cloud will store all their data in the cloud. This means there is no protected health information (PHI) residing on hospital computers, which is a more secure situation than today's current environment.

C. *Workflow challenges Transmission and execution tests*

As it can be difficult to enact change throughout healthcare provider organizations, we may assume that adoption of a cloud model would present significant change management issue for providers. Current processes are often inefficient, relying on paper in many cases to manage patient care. Any transition to a cloud would require significant support from the technology partners to ensure a smooth transition for users. As a part of this workflow transition, serious consideration should be given to staffing needs within the organization's IT department. As the cloud starts to permeate the clinical environment, no longer will the same skill sets be required. Different technology will need to be supported, new training will be required and new skill sets will need to be defined. An organization that had staff working on managing backups and archiving will now migrate to network connections and clinical applications. IT staff will focus on the rollout of the electronic medical record (EMR) instead of managing the storage layer the EMR sits upon. Access to this kind of skill set is in high demand today.

These challenges contribute to slow adoption of cloud technologies but should not stop cloud progress. Organizations are weighing the benefits against the risks. As more providers migrate to the cloud, we will see these challenges overcome with new and innovative solutions.

VII. MODE OF AVAILABILITY

In this work we proposed an architecture for advanced remote data processing in a secure and versatile client–server environment.

To demonstrate the feasibility of the proposed system and the advantages over past studies, we have implemented a medical case study for automatic lesion detection in breast DCE-MRI [24].

The result is a process-on-demand architecture, allowing the radiologist to have access to a secure, versatile and powerful remote CAD system.

The aim of the proposed architecture is the possibility of easily integrating a pre-existing medical image processing software within a complete CAD system deployed on a server machine and shared with many workstations. This has the benefit to allow the physician to not change the user interface he is accustomed to, but to extend the pre-existent software

VIII. SUMMARY

In this work we proposed an architecture for advanced remote data processing in a secure and versatile client–server environment. To demonstrate the feasibility of the proposed system and the advantages over past studies, we have implemented a medical case study for automatic lesion detection in breast DCE-MRI [24]. The result is a process-on-demand architecture, allowing the radiologist to have access to a secure, versatile and powerful remote CAD system.

The aim of the proposed architecture is the possibility of easily integrating a pre-existing medical image processing software within a complete CAD system deployed on a server machine and shared with many workstations. This has the benefit to allow the physician to not change the user interface he is accustomed to, but to extend the pre-existent software

REFERENCES

[1] Buyya R, Yeo CS, Venugopal S (2008) Market-oriented cloud computing: vision, hype, and reality for delivering it services as computing utilities, In: 10th IEEE international conference on high performance computing and communications, 2008. HPCC2008, IEEE, New York, pp 5–13

[2]. Bryant R, Katz RH, Lazowska ED (2008) Big-data computing: creating revolutionary breakthroughs in commerce, science and society

[3]. Majithia S, Taylor I, Shields M, Wang I (2003) Triana as a graphical web services composition toolkit. In: Proceedings of the UK eScience all hands meeting, pp 2–4

[4]. Curcin V, Ghanem M (2008) Scientific workflow systems-can one size fit all? In: Cairo international biomedical engineering conference, 2008. CIBEC 2008. IEEE, New York, pp 1–9

[5]. Altintas I, Berkley C, Jaeger E, Jones M, Ludascher B, Mock S (2004) Kepler: an extensible system for design and execution of scientific workflows. In: Proceedings of the 16th international conference on scientific and statistical database management, 2004. IEEE, New York, pp 423–424

[6]. Rex DE, Ma JQ, Toga AW (2003) The loni pipeline processing environment. Neuroimage 19(3):1033–1048

[7]. Wolstencroft K, Haines R, Fellows D, Williams A, Withers D, Owen S, Soiland-Reyes S, Dunlop I, Nenadic A, Fisher P et al (2013) The taverna workflow suite: designing and executing workflows of web services on the desktop, web or in the cloud. Nucleic Acids Res gkt328

[8]. SvantessonD, ClarkeR(2010) Privacy and consumer risks in cloud computing. Comput Law Secur Rev 26(4):391–397

[9]. Scheinine AL, Donizelli M, Pescosolido M (1998) An objectoriented client–server system for interactive segmentation of medical images using the method of active contours. In: Bildverarbeitung für die Medizin 1998. Springer, New York, pp 308–312

[10]. Mayer A, Meinzer H-P (1999) High performance medical image processing in client/server-environments. Comput Methods Programs Biomed 58(3):207–217

[11]. Yacoub SM,AmmarHH(1999) The development of a client/server